Multiscale Riccati Equations and a Two-Sweep Algorithm
for the Optimal Fusion of Multiresolution Data

K. C. Chou[1]

A. S. Willsky[1]

Abstract

In a previous paper [7] we introduced a class of multiscale dynamic models evolving on dyadic trees in which each level in the tree corresponds to the representation of a signal at a particular scale. One of the estimation algorithms suggested in [7] led to the introduction of a new class of Riccati equations describing the evolution of the estimation error covariance as multiresolution data is fused in a fine-to-coarse direction. This equation can be thought of as having 3 steps in its recursive description: a measurement update step, a fine-to-coarse prediction step, and a **fusion** step. In this paper we analyze this class of equations. In particular by introducing several rudimentary elements of a system theory for processes on trees we develop bounds on the error covariance and use these in analyzing stability and steady-state behavior of the fine-to-coarse filter and the Riccati equations. While this analysis is similar in spirit to that for standard Riccati equations and Kalman filters, there are substantial differences that arise in the multiscale context. For example, the asymmetry of the dyadic tree makes it necessary to define multiscale processes via a coarse-to-fine dynamic model and also to define the first step in a fusion processor in the opposite direction − i.e. fine-to-coarse. Also, the notions of stability, reachability, and observability are different. Most importantly for the analysis here, we will see that the fusion step in the fine-to-coarse filter and Riccati equation requires that we focus attention on the maximum likelihood estimator in order to develop a stability and steady-state theory.

	Report Documentation Page		

1. REPORT DATE **FEB 1990**	2. REPORT TYPE	3. DATES COVERED **00-02-1990 to 00-02-1990**	
4. TITLE AND SUBTITLE **Multiscale Riccati Equations and a Two-Sweep Algorithm for the Optimal Fusion of Multiresolution Data**		5a. CONTRACT NUMBER	
		5b. GRANT NUMBER	
		5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)		5d. PROJECT NUMBER	
		5e. TASK NUMBER	
		5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Massachusetts Institute of Technology,Laboratory for Information and Decision Systems,77 Massachusetts Avenue,Cambridge,MA,02139-4307**		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)	
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**			
13. SUPPLEMENTARY NOTES			
14. ABSTRACT			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **53**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Contents

List of Figures

1 Introduction

Multiscale signal analysis is presently an extremely active research topic due in large part to the emerging theory of wavelet transforms [8,10,11] and to a broad array of applications in which multiresolution analysis seems to be needed or natural. Our work in this area [7,2] has been motivated by a desire to develop multiscale statistical models, inspired by the structure of wavelet transforms, and which could then provide the foundation for statistically optimal multiresolution processing algorithms. In particular in [7] we introduced a class of multiscale state models evolving in a coarse-to-fine direction on a dyadic tree and presented several algorithms for optimal estimation for these processes, i.e. for statistically optimal fusion of multiresolution measurements. In this paper we take a much more careful look at one of these algorithms and develop the required system-theoretic concepts for systems on trees that allow us to analyze and to understand more deeply the structure and properties of this class of multiresolution data fusion algorithms.

In the next section we briefly review the modeling framework introduced in [7] and one of the estimation procedures derived therein. In particular, as we discuss, the wavelet transform makes it natural to define multiscale models evolving from coarse to fine resolution representations. On the other hand, the particular estimation algorithm analyzed here – a two sweep algorithm in the spirit of the Rauch-Tung-Striebel smoothing algorithm – **must** have as its first step a sweep evolving in the opposite direction, i.e. from fine to coarse scales. Furthermore this sweep, which resembles a Kalman filter recursion(although now in scale), has an additional step not found in temporal processing corresponding to the **fusion** of information as we move from fine-to-coarse scales.

The remainder of this paper then analyzes in detail the qualitative properties of this fine-to-coarse filtering step. In particular our main results center on the stability of this step and the convergence to steady-state. As we will see, the fusion step makes it necessary to view the optimal estimator as producing a maximum likelihood(ML) estimate which is then combined with prior statistics, and it is the dynamics of the ML

estimate recursion which must be analyzed. Also, in order to analyze this recursion we need to develop several system-theoretic notions for fine-to-coarse recursions on dyadic trees. In particular, in Section 3 we motivate and define the ML version of our fine-to-coarse Kalman filter. In Section 4 we develop notions of reachability and observability which we then use in Section 5 to obtain bounds on the error covariance of the filter. In Section 6 we then define and analyze l_p-stability for fine-to-coarse recursions. As we will see, the conditions for stability depend strongly on the choice of p. In Section 7 we then use our bounds on the error covariance as the basis for a Lyapunov proof of l_2-stability of the fine-to-coarse filter, while in Section 8 we present results on the steady-state filter.

2 Multiscale Stochastic Processes on Trees and Their Estimation

As described in [10,11], the wavelet transform of a function $f(x)$ provides a sequence of approximations of the signal, at successively finer scales, consisting of linear combinations of shifted versions of a single function $\phi(x)$ compressed or expanded to match the scale in question. That is the approximation of $f(x)$ at the mth scale is given by

$$f_m(x) = \sum_{n=-\infty}^{+\infty} f(m,n)\phi(2^m x - n) \qquad (2.1)$$

As we describe in [7], the evolution of this approximation from scale to scale describes a *dynamical* relationship between the coefficients $f(m,n)$ at one scale and those at the next. Indeed this relationship defines a lattice on the points (m,n), where $(m+1,k)$ is connected to (m,n) if $f(m,n)$ influences $f(m+1,k)$. For example the so-called **Haar approximation**, in which each $f(m,n)$ is simply an average of $f(x)$ over an interval of length 2^{-m}, naturally defines a dyadic tree structure on the points (m,n) in which each point has two equally-weighted descendents corresponding to the two subintervals of length 2^{-m-1} at the $(m+1)$st scale obtained from the corresponding interval of length 2^{-m} at the mth scale.

The preceding development provides the motivation for the study of stochastic processes $x(m,n)$ defined on the types of lattices just described. While we have performed some analysis for the most general of these lattices [6], the work in [7] and in this paper focus on the dyadic tree. Let us make several comments about this case. First, as illustrated in Figure 1, with this and any of the other lattices, the scale index m is time-like. For example it defines a natural direction of recursion for our representation, namely a signal is synthesized via a coarse-to-fine recursion. In the case of our tree, with increasing m - i.e. the direction of synthesis - denoting the forward direction, we then can define a unique backward shift γ^{-1} and two forward shifts α and β(see Figure 1). Also, for notational convenience we denote each node of the tree by a single abstract index t and let T denote the set of all nodes. Thus if $t = (m,n)$ then $\alpha t = (m+1,2n)$, $\beta t = (m+1,2n+1)$, and $\gamma^{-1}t = (m-1,[\frac{n}{2}])$ where
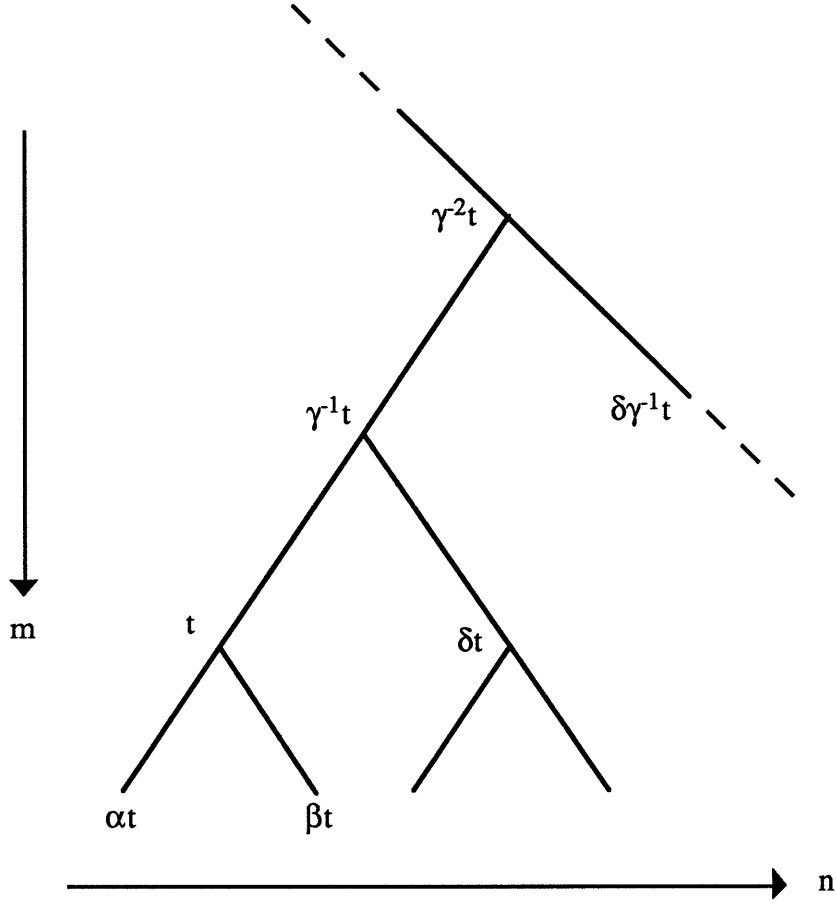
Figure 1: Dyadic Tree Representation

$[x]$ =integer part of x. Also we use the notation $m(t)$ to denote the scale(i.e. the m-component of t). Finally, it is worth noting that while we have described multi-scale representations for continuous-time signals on $(-\infty, \infty)$, they can also be used for signals on compact intervals or in discrete-time. For example a signal defined for $t = 0, 1, ..., 2^{M-1}$ can be represented by M scales, each of which represents in essence an averaged, decimated version of the finer scale immediately below it. In this case the tree of Figure 1 has a bottom level, representing the samples of the signal itself, and a single root node, denoted by 0, at the top. Such a root node also exists in the representation of continuous-time signals defined on a compact interval.

With the preceding as motivation we introduced in [7] the following class of state-space models on trees:

$$x(t) = A(m(t))x(\gamma^{-1}t) + B(m(t))w(t) \tag{2.2}$$

where $\{w(t), t \in T\}$ is a set of independent, zero-mean Gaussian random variables. If we are dealing with a tree with unique root node, 0, we require $w(t)$ to be independent of $x(0)$, the zero-mean initial condition. The covariance of $w(t)$ is I and that of $x(0)$ is $P_x(0)$. If we wish the model eq.(2.2) to define a process over the entire infinite tree, we simply require that $w(t)$ is independent of the "past" of x, i.e. $\{x(\tau)|m(\tau) < m(t)\}$. If $A(m)$ is invertible for all m, this is equivalent to requiring $w(t)$ to be independent of *some* $x(\tau)$ with $\tau \neq t$, $m(\tau) < m(t)$.

Let us make several comments about this model. Note first that the model *does* evolve along the tree, as both $x(\alpha t)$ and $x(\beta t)$ evolve from $x(t)$. Secondly, we note that this process has a Markovian property: given x at scale m, x at scale $m + 1$ is independent of x at scales less than or equal to $m - 1$. Indeed for this to hold all we need is for w to be independent from scale to scale and not necessarily at each individual node. Also while the analysis we perform is easily extended to the case in which A and B are arbitrary functions of t, we have chosen to focus here on a translation-invariant model: we allow these quantities to depend only on scale. As we will see this leads to significant computational efficiencies and also, when this dependence is chosen appropriately, these models lead to processes possessing self-similar properties from scale to scale.

Note that the second-order statistics of $x(t)$ are easily computed. In particular the covariance $P_x(t) = E[x(t)x^T(t)]$ evolves according to a Lyapunov equation on the tree:

$$P_x(t) = A(m(t))P_x(\gamma^{-1}t)A^T(m(t)) + B(m(t))B^T(m(t)) \tag{2.3}$$

Note in particular that if $P_x(\tau)$ depends only on $m(\tau)$ for $m(\tau) \leq m(t) - 1$, then $P_x(t)$ depends only on $m(t)$. We will assume that this is the case and therefore will write $P_x(t) = P_x(m(t))$. Note that this is always true if we are considering the subtree with single root node 0. Also if $A(m)$ is invertible for all m, and if $P_x(t) = P_x(m(t))$ at

some scale(i.e. at all t for which $m(t)$ equals m for some m), then $P_x(t) = P_x(m(t))$ for *all t*. Furthermore, if $A(m(t)) = A$ is stable and if $B(m(t)) = B$, let P_x be the solution to the algebraic Lyapunov equation

$$P_x = AP_xA^T + BB^T \tag{2.4}$$

In this case if $P_x(0) = P_x$(if we have a root node), or if we assume that $P_x(\tau) = P_x$ for $m(\tau)$ sufficiently negative[1], then $P_x(t) = P_x$ for all t, and we have the stationary model.

As we will see in a moment, the multiscale estimation algorithm we will analyze involves a fine-to-coarse recursion requiring a corresponding version of eq.(2.2). Assuming that $A(m)$ is invertible for all m we can directly apply the results of [12]:

$$x(\gamma^{-1}t) = F(m(t))x(t) - A^{-1}(m(t))B(m(t))\tilde{w}(t) \tag{2.5}$$

with

$$
\begin{aligned}
F(m(t)) &= A^{-1}(m(t))[I - B(m(t))B^T(m(t))P_x^{-1}(m(t))] \\
&= P_x(m(t) - 1)A^T(m(t))P_x^{-1}(m(t)) \tag{2.6}
\end{aligned}
$$

and where

$$
\begin{aligned}
\tilde{w}(t) &= w(t) - E[w(t)|x(t)] \tag{2.7} \\
E[\tilde{w}(t)\tilde{w}^T(t)] &= I - B^T(m(t))P_x^{-1}(m(t))B(m(t)) \tag{2.8} \\
&\triangleq \tilde{Q}(m(t))
\end{aligned}
$$

Note that $\tilde{w}(t)$ is a white noise process along all upward paths on the tree – i.e. $\tilde{w}(s)$ and $\tilde{w}(t)$ are uncorrelated if $t = \gamma^{-r}s$ or $s = \gamma^{-r}t$ for some r; otherwise $\tilde{w}(s)$ and $\tilde{w}(t)$ are **not** uncorrelated.

---

[1]Once again if A is invertible, if $P_x(t) = P_x$ at *any* single node, $P_x(t) = P_x$ at *all* nodes.

In [7] we consider the estimation of the stochastic process described by eq.(2.2) based on the measurements

$$y(t) = C(m(t))x(t) + v(t) \tag{2.9}$$

where $\{v(t), t \in T\}$ is a set of independent zero-mean Gaussian random variables independent of $x(0)$ and $\{w(t), t \in T\}$. The covariance of $v(t)$ is $R(m(t))$. The model eq.(2.9) allows us to consider multiple resolution measurements of our process. The single resolution problem, i.e. when $C(m) = 0$ unless $m = M$(the finest level), is also of interest as it corresponds to the problem of restoring a noise corrupted version of a stochastic process possessing a multi-scale description.

Three different algorithm structures are described in [7]. One of these is a generalization of the well-known Rauch-Tung-Striebel(RTS) smoothing algorithm for causal state models. Recall that the standard RTS algorithm involves a forward Kalman filtering sweep followed by a backward sweep to compute the smoothed estimates. The generalization to our models on trees has the same structure, with several important differences. First for the standard RTS algorithm the procedure is completely symmetric with respect to time – i.e. we can start with a reverse-time Kalman filtering sweep followed by a forward smoothing sweep. For processes on trees, the Kalman filtering sweep **must** proceed from fine-to-coarse(i.e. in the reverse direction from that in which the model eq.(2.2) is defined) followed by a coarse-to-fine smoothing sweep[2]. Furthermore the Kalman filtering sweep, using the backward model eq.'s(2.5-2.8) is somewhat more complex for processes on trees. In particular one full step of the Kalman filter recursion involves a measurement update, **two** parallel backward predictions(corresponding to backward prediction along both of the paths descending from a node), and the **fusion** of these predicted estimates. This last step has no counterpart for state models evolving in *time* and is one of the major reasons for the differences between the analysis of temporal Riccati equations and that presented in this paper.

---

[2]The reason for this is not very complex. To allow the measurement on the tree at one point to contribute to the estimate at another point on the same level of the tree, one must use a recursion that first moves up and then down the tree. Reversing the order of these steps does not allow one to realize such contributions.

$\widehat{x}$(t|t) is based on measurements in

$\widehat{x}$(t|t+) is based on
measurements in

t

αt

βt
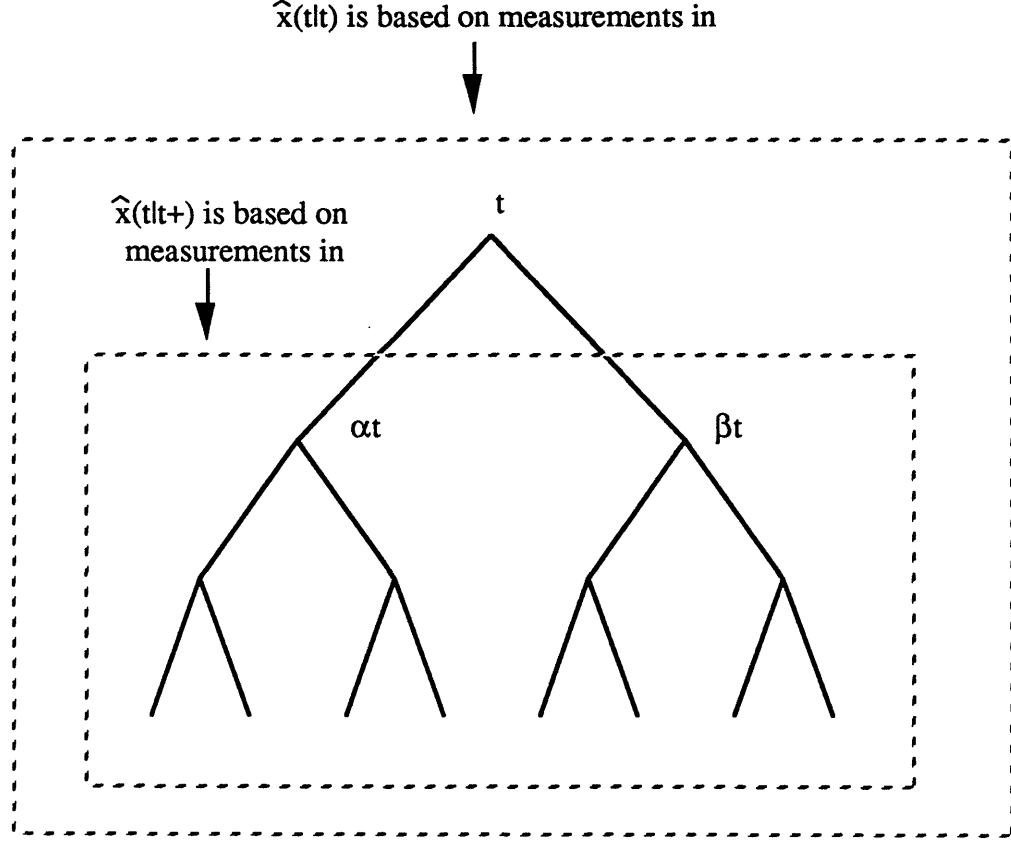
Figure 2: Representation of Meaurement Update and Merged Estimates

To begin let us define some notation:

$$Y_t = \{y(s)|s = t \text{ or } s \text{ is a descendent of } t\}$$

$$= \{y(s)|s \in (\alpha,\beta)^*t , m(s) \le M\} \tag{2.10}$$

$$Y_t^+ = \{y(s)|s \in (\alpha,\beta)^*t , t < m(s) \le M\} \tag{2.11}$$

$$\hat{x}(\cdot|t) = E[x(\cdot)|Y_t] \tag{2.12}$$

$$\hat{x}(\cdot|t+) = E[x(\cdot)|Y_t^+] \tag{2.13}$$

The interpretation of these estimates is provided in Figure 2.

As developed in [7], the Kalman filter and Riccati equation recursions have the following steps. To begin, consider the measurement update. Specifically, suppose that

we have computed $\hat{x}(t|t+)$ and the corresponding error covariance, $P(m(t)|m(t)+)$; the fact that this depends only on scale should be evident from the structure of the problem. Then, standard estimation results yield

$$\hat{x}(t|t) = \hat{x}(t|t+) + K(m(t))[y(t) - C(m(t))\hat{x}(t|t+)] \tag{2.14}$$

$$K(m(t)) = P(m(t)|m(t)+)C^T(m(t))V^{-1}(m(t)) \tag{2.15}$$

$$V(m(t)) = C(m(t))P(m(t)|m(t)+)C^T(m(t)) + R(m(t)) \tag{2.16}$$

and the resulting error covariance is given by

$$P(m(t)|m(t)) = [I - K(m(t))C(m(t))]P(m(t)|m(t)+) \tag{2.17}$$

Note that the computations begin on the finest level(m(t)=M) with $\hat{x}(t|t+) = 0$, $P(M|M+) = P_x(M)$.

Suppose now that we have computed $\hat{x}(\alpha t|\alpha t)$ and $\hat{x}(\beta t|\beta t)$. Note that $Y_{\alpha t}$ and $Y_{\beta t}$ are disjoint and these estimates can be calculated in parallel. Furthermore, once again they have equal error covariances, $P(m(t) + 1|m(t) + 1)$. We then compute $\hat{x}(t|\alpha t)$ and $\hat{x}(t|\beta t)$ which are given by

$$\hat{x}(t|\alpha t) = F(m(t) + 1)\hat{x}(\alpha t|\alpha t) \tag{2.18}$$

$$\hat{x}(t|\beta t) = F(m(t) + 1)\hat{x}(\beta t|\beta t) \tag{2.19}$$

with corresponding identical error covariances given by

$$P(m(t)|m(t) + 1) = F(m(t) + 1)P(m(t) + 1|m(t) + 1)F^T(m(t) + 1) + \mathcal{Q}(m(t) + 1) \tag{2.20}$$

$$\mathcal{Q}(m(t) + 1) = A^{-1}(m(t) + 1)B(m(t) + 1)\tilde{Q}(m(t) + 1)B^T(m(t) + 1)A^{-T}(m(t) + 1) \tag{2.21}$$

These estimates must then be fused to form $\hat{x}(t|t+)$ as follows:

$$\hat{x}(t|t+) = P(m(t)|m(t)+)P^{-1}(m(t)|m(t) + 1)[\hat{x}(t|\alpha t) + \hat{x}(t|\beta t)] \tag{2.22}$$

$$P(m(t)|m(t)+) = [2P^{-1}(m(t)|m(t) + 1) - P_x^{-1}(t)]^{-1} \tag{2.23}$$

The interpretation of these equations is that we are fusing together two estimates based on independent sources of information, namely $Y_{\alpha t}$ and $Y_{\beta t}$, and on one common information source, namely the prior statistics of $x(t)$. Eq.(2.23) ensures that this common information is accounted for only once in the fused estimate.

The analysis in the remainder of this paper focuses on the upward Kalman filtering sweep. For completeness we describe the subsequent downward smoothing sweep. Specifically, when we reach the top node of the tree, the resulting updated estimate is the smoothed estimate at that point which then serves as the initial condition for the downward recursion along the tree. This recursion combines the smoothed estimate $\hat{x}_s(\gamma^{-1}t)$ with the filtered estimates from the upward sweep to produce $\hat{x}_s(t)$:

$$\hat{x}_s(t) = \hat{x}(t|t) + P(m(t)|m(t))F^T(m(t))P^{-1}(m(t) - 1|m(t))\left[\hat{x}_s(\gamma^{-1}t) - \hat{x}(\gamma^{-1}t|t)\right]$$

$$(2.24)$$

3   Maximum Likelihood Estimator

In this section we examine the difficulties in analyzing our filtering equations. These difficulties point to the need to decompose the filter into two parts; one representing the filter initialized with no prior information, the ML filter, and the other representing our estimate of the mean of the process.

We rewrite the set of Riccati equations for our filtering problem as follows.

$$
\begin{aligned}
P(m|m+1) &= F(m+1)P(m+1|m+1)F^T(m+1) \\
&\quad + G(m+1)\tilde{Q}(m+1)G^T(m+1) \tag{3.1} \\
P^{-1}(m|m) &= P^{-1}(m|m^+) + C^T(m)R^{-1}(m)C(m) \tag{3.2} \\
P^{-1}(m|m^+) &= 2P^{-1}(m|m+1) - P_x^{-1}(m) \tag{3.3}
\end{aligned}
$$

where

$$
G(m(t)) \triangleq -A^{-1}(m(t))B(m(t)) \tag{3.4}
$$

Note that we can combine eq.(3.2,3.3) into the following single equation.

$$
\begin{aligned}
P^{-1}(m|m) &= 2P^{-1}(m|m+1) - P_x^{-1}(m) + C^T(m)R^{-1}(m)C(m) \\
&= P^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \\
&\quad + P^{-1}(m|m+1) - P_x^{-1}(m) \tag{3.5}
\end{aligned}
$$

The Riccati equations for our optimal filter, eq.'s(3.1-3.3), differ from standard Riccati equations in two respects: 1) the explicit presence of the prior state covariance $P_x(m(t))$ and 2) the presence of a scaling factor of 2 in eq.(3.3). The scaling factor is intrinsic to our Riccati equations and is due to the fact that we are fusing **pairs** of parallel information paths in going from level to level. The presence of $P_x(m(t))$ in the Riccati equations accounts for the double counting of prior information in performing this merge.

The presence of this term points to a significant complication in analyzing this filter. Specifically, in standard Kalman filtering analysis the Riccati equation for the error covariance can be viewed simply as the covariance of the error equations,

which can be analyzed directly without explicitly examining the state dynamics since the error evolves as a state process itself. This is apparently **not** the case here because of the explicit presence of $P_x(m)$ in eq.(3.5). Indeed as we show later in this section, if one examines the backward model eq.'s(2.5-2.8) and the Kalman filter eq.'s(2.14,2.18,2.19,2.22) one finds that the upward dynamics for the error $x(t) - \hat{x}(t|t)$ are **not** decoupled from $x(t)$ **unless** $P_x^{-1}(m(t)) = 0$. This motivates the following decomposition of the estimator into a dynamic part based on $P_x^{-1} = 0$(the ML estimator) followed by a gain adjustment to account for prior information.

To be precise, let $P_{ML}(m|m+1)$ and $P_{ML}(m|m)$ denote the estimates produced by our upward Kalman filter assuming that $P_x^{-1}(m) = 0$. These satisfy the following Riccati equation, which doesn't depend explicitly on $P_x(m)$.

$$P_{ML}(m|m+1) = F(m+1)P_{ML}(m+1|m+1)F^T(m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1)$$

$$(3.6)$$

$$P_{ML}^{-1}(m|m) = 2P_{ML}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \qquad (3.7)$$

Note that the filtering equations for the ML estimator correspond exactly with the equations for the optimal(Bayesian) filter with $P_{ML}(m|m)$ and $P_{ML}(m|m+1)$ being substituted for $P(m|m)$ and $P(m|m+1)$. We refer to these as the **ML filtering equations**.

Before elaborating further on the ML estimator, we describe its relationship to the optimal estimator. The two are related in the following way.

$$\hat{x}(t|t) = P(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t))\hat{x}_{ML}(t|t) \qquad (3.8)$$

$$P^{-1}(m(t)|m(t)) = P_{ML}^{-1}(m(t)|m(t)) + P_x^{-1}(m(t)) \qquad (3.9)$$

To derive these relationships we start by writing

$$Y_t = \mathcal{H}_t x(t) + \theta(t) \qquad (3.10)$$

where

$$E[\theta(t)x^T(t)] = 0 \qquad (3.11)$$

$$E[\theta(t)\theta^T(t)] = \mathcal{R}_t \qquad (3.12)$$

Recall that Y_t is the set $\{y(s)|s = t$ or s is a descendent of $t\}$. Eq.(3.10) follows directly from our downward model for the process $x(t)$ on the tree. From eq.(3.10) we can write the following maximum likelihood estimate.

$$\hat{x}_{ML}(t|t) = (\mathcal{H}_t^T \mathcal{R}_t^{-1} \mathcal{H}_t) \mathcal{H}_t^T \mathcal{R}_t^{-1} Y_t \qquad (3.13)$$

$$P_{ML}^{-1}(m(t)|m(t)) = \mathcal{H}_t^T \mathcal{R}_t^{-1} \mathcal{H}_t \qquad (3.14)$$

Note that $\hat{x}_{ML}(t|t)$ can be computed using ML filtering equations. This is true since the ML filter computes the best estimate in the sense of minimizing the mean-square error given no initial prior information, which from the invertibility of $F(m)$ and from our Lyapunov equation for the evolution of the state covariance is **equivalent** to the best estimate at some point t given $P_x^{-1}(m(t)) = 0$. Furthermore, since $\theta(t)$ is uncorrelated with $x(t)$ we can write the Bayesian estimate as follows.

$$\hat{x}(t|t) = P^{-1}(m(t)|m(t))(P_{ML}^{-1}(m(t)|m(t))\hat{x}_{ML}(t|t) + P_x^{-1}(t)m(x(t)))$$

$$(3.15)$$

$$P^{-1}(m(t)|m(t)) = P_{ML}^{-1}(m(t)|m(t)) + P_x^{-1}(m(t)) \qquad (3.16)$$

where $m(x(t)))$ is the mean of $x(t)$. But since we consider $x(t)$ to be a zero-mean process eq.(3.15) and eq.(3.8) are equivalent.

There are several reasons for viewing the optimal estimator in this way. One is that the ML Riccati equations are simpler because they do not include the explicit presence of the prior information $P_x^{-1}(m(t))$. This simplicity is significant in that the ML Riccati equations are readily amenable to stability analysis. The important reason mentioned previously for focusing our analysis on the ML filter, and perhaps a deeper one, is that the error dynamics for the optimal filter cannot be written as a noise driven process with closed-loop dynamics whereas the error dynamics for the ML filter **can**. Let us flesh out this last point in more detail.

Let us begin by examining the dynamics of our filter in the upward sweep of the RTS algorithm, eq.'s(2.14-2.17, 2.18-2.21,2.22,2.23). We can rewrite the dynamics of the filter in update form, eq.(2.14), as follows.

$$\hat{x}(t|t) = L(m(t))F(m(t) + 1)(\hat{x}(\alpha t|\alpha t) + \hat{x}(\beta t|\beta t))$$

$$+ \quad K(m(t))y(t) \tag{3.17}$$

$$L(m(t)) \quad = \quad P(m(t)|m(t))P^{-1}(m(t)|m(t)+1) \tag{3.18}$$

We can also write the dynamics for our process in a similarly symmetric form.

$$x(t) = \frac{1}{2}F(m(t)+1)[x(\alpha t) + x(\beta t)] + \frac{1}{2}G(m(t)+1)[\tilde{w}(\alpha t) + \tilde{w}(\beta t)] \tag{3.19}$$

We can easily rewrite eq.(3.17) as

$$\hat{x}(t|t) \quad = \quad (I - K(m(t))C(m(t)))L'(m(t))F(m(t)+1)(\hat{x}(\alpha t|\alpha t)$$
$$+ \quad \hat{x}(\beta t|\beta t)) + K(m(t))y(t) \tag{3.20}$$

$$L'(m(t)) \quad = \quad P(m(t)|m(t)+)P^{-1}(m(t)|m(t)+1) \tag{3.21}$$

By doing straightforward manipulations on eq.(3.20) and eq.(3.19) we can get

$$\tilde{x}(t|t) \quad = \quad (I - K(m(t))C(m(t)))x(t) - K(m(t))v(t)$$
$$- \quad (I - K(m(t))C(m(t)))L'(m(t))F(m(t)+1)(\hat{x}(\alpha t|\alpha t) + \hat{x}(\beta t|\beta t))$$
$$\tag{3.22}$$

$$\tilde{x}(t|t) \quad = \quad x(t) - \hat{x}(t|t) \tag{3.23}$$

The difficulty in proceeding any further with eq.(3.22) lies in the presence of the term $L'(m(t))$. In standard filtering $L'(m(t)) = I$; said another way there is no difference between $P(m(t)|m(t)+)$ and $P(m(t)|m(t)+1)$. Let us write down the equations for the ML filter and its corresponding error.

$$\hat{x}_{ML}(t|t) \quad = \quad \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))F(m(t)+1)(\hat{x}_{ML}(\alpha t|\alpha t) + \hat{x}_{ML}(\beta t|\beta t))$$
$$+ \quad K_{ML}(m(t))y(t) \tag{3.24}$$

$$\tilde{x}_{ML}(t|t) \quad = \quad (I - K_{ML}(m(t))C(m(t)))x(t) - K_{ML}(m(t))v(t)$$
$$- \quad \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))F(m(t)+1)(\hat{x}_{ML}(\alpha t|\alpha t) + \hat{x}_{ML}(\beta t|\beta t)$$
$$\tag{3.25}$$

By substituting eq.(3.19) into eq.(3.25) we get

$$\tilde{x}_{ML}(t|t) = \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))F(m(t)+1)(\tilde{x}_{ML}(\alpha t|\alpha t) + \tilde{x}_{ML}(\beta t|\beta t)$$
$$+ \frac{1}{2}(I - K_{ML}(m(t))C(m(t)))G(m(t)+1)(\tilde{w}(\alpha t) + \tilde{w}(\beta t)) - K_{ML}(m(t))v(t)$$

$$(3.26)$$

Note that eq.(3.26) has the same algebraic structure as the the equations for the error dynamics of the standard Kalman filter except for the scaling factor of $\frac{1}{2}$ and the fact that there are two terms in the immediate past being merged. Both the scaling factor and the merging of pairs of points is crucial to the study of the stability of the filter. As we will see in Section 7 the appropriate scaling factor is necessary for controlling in some sense the potential growth that might occur in merging points.

Also, for future reference, let us rewrite eq.(3.26) using the following equality:

$$\frac{1}{2}(I - K_{ML}(m(t))C(m(t))) = P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t)+1) \qquad (3.27)$$

We can rewrite eq.(3.26) as

$$\tilde{x}_{ML}(t|t) = P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t)+1)F(m(t)+1)(\tilde{x}_{ML}(\alpha t|\alpha t) + \tilde{x}_{ML}(\beta t|\beta t)$$
$$+ P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t)+1)G(m(t)+1)(\tilde{w}(\alpha t) + \tilde{w}(\beta t)) - K_{ML}(m(t))v(t)$$

$$(3.28)$$

4   Reachability, Observability, and Reconstructibility

In this section we develop certain system theoretic constructs which are useful in analyzing both the stability and the steady-state characteristics of our filter. In particular we define notions of reachability, observability, and reconstructibility on dyadic trees in terms of system dynamics going up the tree.

4.1   Upward Reachability

We begin with the notion of reachability for a system defined going up a tree. Analogous to the standard time-series case, reachability involves the notion of being able to reach arbitrary states at some point t on the tree given arbitrary inputs in the past where in the case of processes evolving up a tree the past refers to points in the subtree under t. Recall that we can rewrite the dynamics for our backward process up the tree, eq.(2.5), in the following form.

$$x(t) = \frac{1}{2}F(m(t)+1)[x(\alpha t) + x(\beta t)] + \frac{1}{2}G(m(t)+1)[\tilde{w}(\alpha t) + \tilde{w}(\beta t)] \qquad (4.1)$$

Also, recall that in our backward model $\tilde{w}(t)$ is a white noise process along upward paths on the tree. For the analysis of reachability, however, we simply view $\tilde{w}(t)$ as the input to the system eq.(4.1).

We define the following vectors,

$$X_{M,t_0} \triangleq [x^T(\alpha^M t_0), x^T(\beta\alpha^{M-1}t_0), ...x^T(\beta^M t_0)]^T \qquad (4.2)$$

$$\tilde{W}_{M,t_0} \triangleq [\ \tilde{w}^T(\alpha t_0) \quad \tilde{w}^T(\beta t_0) \quad ... \quad \tilde{w}^T(\alpha^M t_0)...\tilde{w}^T(\beta^M t_0) \ ]^T \qquad (4.3)$$

which have the following interpretation. Consider an arbitrary point on the tree, t_0. The vector X_{M,t_0} denotes the vector of 2^M points at the Mth level down in the subtree under t_0; i.e. X_{M,t_0} includes all of the nodes at this level that influence the value of $x(t_0)$. The vector $\tilde{W}_{M,t_0}$ comprises the full set of inputs that influences $x(t_0)$ starting from initial condition X_{M,t_0}, i.e. the $\tilde{w}(t)$, in the entire subtree down to M levels from t_0. We define **upward reachability** to be the following.

Definition 4.1 *The system is* **upward reachable** *from X_{M,t_0} to $x(t_0)$ if given any $\overline{X}_{M,t_0}$ and any desired $\overline{x}(t_0)$, it is possible to specify $\tilde{W}_{M,t_0}$ so if $X_{M,t_0} = \overline{X}_{M,t_0}$, then $x(t_0) = \overline{x}(t_0)$.*

In studying conditions for reachability since we are given X_{M,t_0}, we can set it equal to zero without loss of generality. Note that if $X_{M,t_0} = 0$, then we have

$$x(t_0) = \mathcal{G}\tilde{W}_{M,t_0} \tag{4.4}$$

where

$$\mathcal{G} \;\triangleq\; [\; \Psi(0)\;\; \Psi(0)\;\; \Psi(1)\;\; \Psi(1)\;\; \Psi(1)\;\; \Psi(1)\;\; ... \tag{4.5}$$

$$\underbrace{\Psi(M-2)...\Psi(M-2)}_{2^{M-1}\text{ times}} \;\; \underbrace{\Psi(M-1)...\Psi(M-1)}_{2^{M}\text{ times}} \;]$$

$$\Psi(i) \;\triangleq\; (\tfrac{1}{2})^{i+1}\phi(m(t_0), m(t_0)+i)G(m(t_0)+i+1) \tag{4.6}$$

$$\phi(m_1, m_2) \;\triangleq\; \begin{cases} I & m_1 = m_2 \\ F(m_1+1)\phi(m_1+1, m_2) & m_1 < m_2 \end{cases} \tag{4.7}$$

$$\phi(m-1, m) \;\triangleq\; F(m) \tag{4.8}$$

Let us also define the following quantity.

Definition 4.2 Upward-reachability Grammian

$$\mathcal{R}(t_0, M) \;\triangleq\; \mathcal{G}\mathcal{G}^T$$

$$= \sum_{i=0}^{M-1} 2^{-i-1}\phi(m(t_0), m(t_0)+i)G(m(t_0)+i+1)$$

$$\times \;\; G^T(m(t_0)+i+1)\phi^T(m(t_0), m(t_0)+i) \tag{4.9}$$

From eq.(4.4) we see that the ability to reach all possible values of $x(t_0)$ given arbitrary inputs, $\tilde{W}_{M,t_0}$, depends on the rank of the matrix $\mathcal{G}$. This, along with the fact that the rank of $\mathcal{G}$ equals the rank of $\mathcal{G}\mathcal{G}^T$, leads to the following, where $x(t)$ is an n-dimensional vector:

Proposition 4.1 *The system is* **upward reachable** *from X_{M,t_0} to $x(t_0)$ iff $\mathcal{G}$ has rank n iff $\mathcal{R}(t_0, M)$ has rank n.*

Note that $\mathcal{R}(t_0, M)$ bears a strong similarity to the standard reachability grammian for the following system.

$$x(m) = \frac{1}{2}F(m+1)x(m+1) + \frac{1}{2}G(m+1)u(m+1) \qquad (4.10)$$

where the reachability grammian in this case is

$$
\begin{aligned}
\mathcal{R}^*(m, m+M) \;\; &\triangleq\; \sum_{i=0}^{M-1} 2^{-2i-2}\phi(m, m+i)G(m+i+1) \\
&\times\; G^T(m+i+1)\phi^T(m, m+i) \\
&=\; \mathcal{G}^*(\mathcal{G}^*)^T \\
\mathcal{G}^* \;\; &\triangleq\; [\;\; \Psi(0) \quad \Psi(1) \quad ... \quad \Psi(M-2) \quad \Psi(M-1)\;\;] \qquad (4.11)
\end{aligned}
$$

In fact it is evident from the definitions in eq.'s(4.5,4.11) that the rank of $\mathcal{G}$ is equivalent to the rank of $\mathcal{G}^*$. This leads to the following corollary.

Corollary 4.1 *The system is **upward reachable** from X_{M,t_0} to $x(t_0)$ iff for any $\alpha, \beta \neq 0$ $\mathcal{R}^*_{\alpha,\beta}(m(t_0), m(t_0) + M)$ has rank n, where $\mathcal{R}^*_{\alpha,\beta}(m(t_0), m(t_0) + M)$ is the reachability grammian for the system*

$$x(m) = \alpha F(m+1)x(m+1) + \beta G(m+1)u(m+1) \qquad (4.12)$$

Note that if F and G are constant in eq.(4.1), then reachability is equivalent to the usual condition, i.e. $rank[G|FG|...|F^{M-1}G] = n$.

4.2 Upward Observability and Reconstructibility

We develop the notion of observability and the notion of reconstructibility on trees. Defined on trees, observability corresponds to the notion of being able to uniquely determine the points at the bottom of a subtree, i.e. the "initial conditions", given knowledge of the inputs and observations in the subtree. It is also useful to develop the weaker notion corresponding to being able to uniquely determine the single point at the top of a subtree given knowledge of the inputs and observations in the subtree. This notion is analogous to reconstructibility for standard systems; thus, we adopt the same term for the notion on trees.

Let us define

$$Y_{M,t_0} \triangleq [\ y^T(t_0)|\quad y^T(\alpha t_0),\quad y^T(\beta t_0)|\quad ...\quad |y^T(\alpha^M t_0), ... y^T(\beta^M t_0)\ ]^T \qquad (4.13)$$

where

$$y(t) = C(m(t))x(t) \qquad (4.14)$$

Definition 4.3 *The system is* **upward observable** *from X_{M,t_0} to $x(t_0)$ if given knowledge of $\tilde{W}_{M,t_0}$ and Y_{M,t_0}, we can uniquely determine X_{M,t_0}.*

Note that if $\tilde{W}_{M,t_0} = 0$ then

$$Y_{M,t_0} = \mathcal{H}_M X_{M,t_0} \qquad (4.15)$$

where $\mathcal{H}_M$ is most easily visualized if we partition it compatibly with the levels of

the observations in Y_{M,t_0}:

$$
\mathcal{H}_M \triangleq
\left[
\begin{array}{ccccccccccccc}
\multicolumn{13}{c}{\overbrace{\hspace{8cm}}^{2^M \text{ blocks}}}\\
\Theta(0) & \Theta(0) & \ldots & & & & & & & & & \ldots & \Theta(0)\\[4pt]
\hline
\Theta(1) & \ldots & & & & \ldots & \Theta(1) & 0 & \ldots & & & \ldots & 0\\
0 & \ldots & & & & \ldots & 0 & \Theta(1) & \ldots & & & \ldots & \Theta(1)\\[4pt]
\hline
\Theta(2) & \ldots & \Theta(2) & 0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & & 0\\
0 & \ldots & 0 & \Theta(2) & \ldots & \Theta(2) & 0 & \ldots & 0 & 0 & \ldots & & 0\\
0 & \ldots & 0 & 0 & \ldots & 0 & \Theta(2) & \ldots & \Theta(2) & 0 & \ldots & & 0\\
0 & \ldots & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 & \Theta(2) & \ldots & & \Theta(2)\\[4pt]
\hline
& & & & \vdots & & & \vdots & & & & &\\
& & & & \vdots & & & \vdots & & & & &\\[4pt]
\hline
\Theta(M) & 0 & \ldots & & & & & & & & & \ldots & 0\\
0 & \Theta(M) & \ldots & & & & & & & & & \ldots & 0\\
\vdots & & & & & & & & & & & & \vdots\\
0 & 0 & \ldots & & & & & & & & & \ldots & \Theta(M)
\end{array}
\right]
$$

$$(4.16)$$

Here

$$
\Theta(i) \triangleq (\frac{1}{2})^{M-i} C(m(t_0)+i)\phi(m(t_0)+i, m(t_0)+M) \qquad (4.17)
$$

As a simple example to help clarify the structure of the matrix $\mathcal{H}_M$ consider the

matrix $\mathcal{H}_2$ for the scale-invariant case, i.e. where $F(m) = F$, $C(m) = C$.

$$\mathcal{H}_2 = \begin{bmatrix} \frac{1}{4}CF^2 & \frac{1}{4}CF^2 & \frac{1}{4}CF^2 & \frac{1}{4}CF^2 \\ \frac{1}{2}CF & \frac{1}{2}CF & 0 & 0 \\ 0 & 0 & \frac{1}{2}CF & \frac{1}{2}CF \\ C & 0 & 0 & 0 \\ 0 & C & 0 & 0 \\ 0 & 0 & C & 0 \\ 0 & 0 & 0 & C \end{bmatrix} \tag{4.18}$$

That is, at level i, there are 2^i measurements each of which provides information about the **sum** of a block of 2^{M-i} components of X_{M,t_0}. Note that this makes clear that upward observability is indeed a very strong condition. Specifically, since successively larger blocks of X_{M,t_0} are summed as we move up the tree, subsequent measurements provide **no** information about the differences among the values that have been summed. For example consider $M = 1$. In this case $y(t)$ contains information about the sum $x(\alpha t) + x(\beta t)$, and thus information about $x(\alpha t) - x(\beta t)$ must come from $y(\alpha t)$ and $y(\beta t)$. This places severe constraints on the system matrices. In particular a necessary condition for observability is that y have dimension larger than $\frac{n}{2}$(otherwise $\mathcal{H}_M$ has fewer rows than columns).

We also define the following.

Definition 4.4 Upward-observability Grammian

$$\mathcal{M}_M \triangleq \mathcal{H}_M^T \mathcal{H}_M \tag{4.19}$$

where

$$\mathcal{M}_k = U(k, 0) \tag{4.20}$$

$$U(k, k) \triangleq \sum_{i=0}^{k} (\frac{1}{2})^{2(k-i)} \phi^T(m(t_0) + i, m(t_0) + k)\mathcal{C}(m(t_0) + i)\phi(m(t_0) + i, m(t_0) + k) \tag{4.21}$$

$$\mathcal{C}(k) \triangleq C^T(k)C(k) \tag{4.22}$$

$$U(k, l) \triangleq \begin{bmatrix} U(k, l+1) & S(k, l) \\ S(k, l) & U(k, l+1) \end{bmatrix} \tag{4.23}$$

and $S(k,l)$ is a block matrix with $2^{k-l-1} \times 2^{k-l-1}$ blocks each of which equals

$$T(k,l) = \sum_{i=0}^{l} (\frac{1}{2})^{2(k-i)} \phi^T(m(t_0)+i, m(t_0)+k) C^T(m(t_0)+i) C(m(t_0)+i) \phi(m(t_0)+i, m(t_0)+k)$$

$$(4.24)$$

Once again we consider the scale-invariant case, this time in order to make explicit the structure of the matrix $\mathcal{M}_M$. The following is $\mathcal{M}_2$ for the scale-invariant case.

$$\mathcal{M}_2 = \begin{bmatrix} M_1 & M_2 & M_3 & M_3 \\ M_2 & M_1 & M_3 & M_3 \\ M_3 & M_3 & M_1 & M_2 \\ M_3 & M_3 & M_2 & M_1 \end{bmatrix} \qquad (4.25)$$

where

$$M_1 = \frac{1}{16} F^{2T} C^T C F^2 + \frac{1}{4} F C^T C F + C^T C \qquad (4.26)$$

$$M_2 = \frac{1}{16} F^{2T} C^T C F^2 + \frac{1}{4} F C^T C F \qquad (4.27)$$

$$M_3 = \frac{1}{16} F^{2T} C^T C F^2 \qquad (4.28)$$

From eq.(4.15) we see that being able to uniquely determine X_{M,t_0} from Y_{M,t_0} is equivalent to requiring the null space of the matrix $\mathcal{H}_M$ to be 0. This leads to the following.

Proposition 4.2 *The system is* **upward observable** *from X_{M,t_0} to $x(t_0)$ iff $\mathcal{N}(\mathcal{H}_M) = 0$ iff $\mathcal{M}_M$ is invertible.*

A much weaker notion than that of observability is the notion of reconstructibility. Reconstructibility requires only the ability to determine the single point at the top of a subtree given knowledge of the inputs and observations in the subtree.

Definition 4.5 *The system is* **upward reconstructible** *from X_{M,t_0} to $x(t_0)$ if given knowledge of $\tilde{W}_{M,t_0}$ and Y_{M,t_0}, we can uniquely determine $x(t_0)$.*

We also define the following.

Definition 4.6 Upward-reconstructibility Grammian

$$
\begin{aligned}
\mathcal{O}(t_0, M) &= I_M \mathcal{H}_M^T \mathcal{H}_M I_M^T \\
&= \sum_{i=0}^{M} 2^i \phi^T(m(t_0) + i, m(t_0) + M) C^T(m(t_0) + i) \\
&\times \quad C(m(t_0) + i)\phi(m(t_0) + i, m(t_0) + M)
\end{aligned}
\tag{4.29}
$$

where

$$
I_M = \underbrace{[I|I|...|I]}_{2^M \ times}
\tag{4.30}
$$

and each I is an n × n identity matrix.

Note that if $\tilde{W}_{M,t_0} = 0$, then

$$
x(t_0) = \Phi(t_0) X_{M,t_0}
\tag{4.31}
$$

where

$$
\Phi(t_0) \triangleq (\frac{1}{2})^M \phi(m(t_0), m(t_0) + M) I_M
\tag{4.32}
$$

Since the condition of reconstructibility only requires being able to uniquely determine the single point $x(t_0)$ from the measurements in the subtree, we guarantee this condition by requiring that any vector in the nullspace, $\mathcal{N}(\mathcal{H}_M)$, is also in the nullspace, $\mathcal{N}(\Phi(t_0))$. We thus have the following, the proof of which can be found in the appendix.

Theorem 4.1 *The system is* **upward reconstructible** *iff $\mathcal{N}(\mathcal{H}) \subseteq \mathcal{N}(\Phi(t_0))$. If $F(m)$ is invertible for all m, this is equivalent to the invertibility of $\mathcal{O}(t_0, M)$.*

Note that $\mathcal{O}(t_0, M)$ bears a strong similarity to the standard **observability** grammian for the following system.

$$
\begin{aligned}
x(m) &= \alpha F(m+1)x(m+1) + G(m+1)u(m+1) \tag{4.33} \\
y(m) &= \beta C(m)x(m) \tag{4.34}
\end{aligned}
$$

where the observability grammian in this case is

$$
\begin{aligned}
\mathcal{O}_{\alpha,\beta}(m(t_0), m(t_0) + M) &\triangleq \sum_{i=0}^{M} \alpha^{2(M-i)}\beta^2 \phi^T(m(t_0) + i, m(t_0) + M) C^T(m(t_0) + i) \\
&\times \quad C(m(t_0) + i)\phi(m(t_0) + i, m(t_0) + M)
\end{aligned}
\tag{4.35}
$$

Corollary 4.2 *Assuming that $F(m)$ is invertible for all m, the system is* **upward** **reconstructible** *from X_{M,t_0} to $x(t_0)$ iff $\mathcal{O}_{\alpha,\beta}(m(t_0), m(t_0) + M)$ has rank n.*

As a final note, let us comment on some similarities and differences between these concepts and results and those for standard temporal systems. First, for standard systems observability implies reconstructibility and the two concepts are equivalent if the state transition matrix is invertible. In our case, observability certainly implies reconstructibility, but the former remains a much stronger condition even if ϕ is invertible. In this case reconstructibility is equivalent to being able to determine the **average values** of the components of the initial state [6]. Note that in contrast our reachability concept going up the tree is actually rather weak since we have **many** control inputs in the subtree to achieve a **single** final state $x(t_0)$. As one might expect there is a dual theory for systems defined moving down the tree, but the tree asymmetry leads to some important differences. In particular, weak and strong concepts are interchanged. For example, observability is concerned with determining the **single** initial state given observations in the subtree under t_0, while reconstructibility corresponds to determining the **entire** vector X_{M,t_0}. In this case if ϕ is invertible observability is equivalent to determining the **average value** of X_{M,t_0}. Similarly, reachability is concerned with reaching arbitrary values for the entire vector X_{M,t_0}, an extremely strong condition. A natural and much weaker condition is achieving an arbitrary average value for X_{M,t_0}. A complete picture of this system theory will be given in [6].

5 Bounds on the Error Covariance of the Filter

In the following sections we will analyze the stability of our upward Kalman filter via Lyapunov methods. As we will see our analysis of the ML filter will require bounds on $P_{ML}(m|m)$, and it will also be necessary to have bounds on $P(m|m)$ in order to infer stability of the optimal filter. Thus, in this section we begin by deriving strict upper and lower bounds for the optimal filter error covariance $P(m|m)$. We then use analogous arguments to derive upper and lower bounds for the ML filter error covariance $P_{ML}(m|m)$. Existence of these bounds depends on conditions that can be expressed in terms of the notions of upward reachability and upward reconstructibility developed in the previous section.

Recall our system whose dynamics are described by eq.(4.1) and whose measurements are described by eq.(4.14). We define the stochastic reachability grammian for this system as follows.

Definition 5.1 Stochastic Reachability Grammian

$$\overline{\mathcal{R}}(t_0, M) \; \triangleq \; \sum_{i=0}^{M-1} 2^{-i-1} \phi(m(t_0), m(t_0) + i) G(m(t_0) + i + 1)$$
$$\times \; \tilde{Q}(m(t_0) + i + 1) G^T(m(t_0) + i + 1) \phi^T(m(t_0), m(t_0) + i) \quad (5.1)$$

We define the stochastic reconstructibility grammian for this system as follows.

Definition 5.2 Stochastic Reconstructibility Grammian

$$\overline{\mathcal{O}}(t_0, M) \; \triangleq \; \sum_{i=0}^{M} 2^i \phi^T(m(t_0) + i, m(t_0) + M) C^T(m(t_0) + i)$$
$$\times \; R^{-1}(m(t_0) + i) C(m(t_0) + i) \phi(m(t_0) + i, m(t_0) + M) \quad (5.2)$$

Among the assumptions that we make under which we prove our bounds is that the matrices $F(m)$, $F^{-1}(m)$, $G(m)$, $\tilde{Q}(m)$, $C(m)$, $R(m)$, and $R^{-1}(m)$ are bounded functions of m. In terms of our reachability and reconstructibility grammians these assumptions mean that for any $M_0 > 0$ we can find $\alpha, \beta > 0$ so that

$$\overline{\mathcal{R}}(t, M_0) \; \leq \; \alpha I \text{ for all } t \quad (5.3)$$

$$\overline{\mathcal{O}}(t, M_0) \; \leq \; \beta I \text{ for all } t \quad (5.4)$$

We define the notion of uniform reachability as follows.

Definition 5.3 *An upward system is* **uniformly reachable** *if there exists* $\gamma, M_0 > 0$ *so that*

$$\overline{\mathcal{R}}(t, M_0) \geq \gamma I \text{ for all } t \tag{5.5}$$

This property insures that the process noise contributes a steady stream of uncertainty into the state. Intuitively, we would expect in this case that the error covariance $P(m|m)$ would never become equal to zero. In fact we prove that under uniform reachabilty $P(m|m)$ is lower bounded by a positive definite matrix.

We also need the notion of uniform reconstructibility, which is formulated as follows.

Definition 5.4 *An upward system is* **uniformly reconstructible** *if there exists* $\delta, M_0 > 0$ *so that*

$$\overline{\mathcal{O}}(t, M_0) \geq \delta I \text{ for all } t \tag{5.6}$$

where M is the bottom level of a tree.

This property insures a steady flow of information about the state of the system. Intuitively, we would expect that under this condition the uncertainty in our estimate remains bounded. In fact we prove that under the condition of uniform reconstructibility the error covariance, $P(m|m)$, is upper bounded.

Without loss of generality we can take M_0 to be the same in eq.'s(5.3-5.6) for any system which is uniformly reachable and reconstructible.

5.1   Upper Bound

We begin by deriving an upper bound for the optimal filter error covariance, $P(m|m)$. The general idea in deriving this bound is to make a careful comparison between the Riccati equations for our optimal filter and the Riccati equations for the standard Kalman filter. First consider the following lemma.

Lemma 5.1 *Given the Riccati equation*

$$
\begin{aligned}
P(m|m+1) &= F(m+1)P(m+1|m+1)F^T(m+1) \\
&+ G(m+1)\tilde{Q}(m+1)G^T(m+1) & (5.7)
\end{aligned}
$$

$$
\begin{aligned}
P^{-1}(m|m) &= P^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \\
&+ P^{-1}(m|m+1) - P_x^{-1}(m) & (5.8)
\end{aligned}
$$

and the Riccati equation

$$
\begin{aligned}
\overline{P}(m|m+1) &= F(m+1)\overline{P}(m+1|m+1)F^T(m+1) \\
&+ G(m+1)\tilde{Q}(m+1)G^T(m+1) & (5.9)
\end{aligned}
$$

$$
\overline{P}^{-1}(m|m) = \overline{P}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \tag{5.10}
$$

we have that

$$
\overline{P}^{-1}(m|m) \le P^{-1}(m|m) \tag{5.11}
$$

Proof

We first note that eq.(5.8) can be rewritten as

$$
P^{-1}(m|m) = P^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) + D^T(m)D(m) \tag{5.12}
$$

where $D^T(m)D(m)$ is positive semi-definite. This follows from the fact that $P(m|m+1) \le P_x(m)$ or $P^{-1}(m|m+1) - P_x^{-1}(m) \ge 0$. The Riccati equation, eq.'s(5.9,5.10), characterizes the error covariance for the optimal filter corresponding to the following filtering problem.

$$
x(m) = F(m+1)x(m+1) + G(m+1)w(m+1) \tag{5.13}
$$

$$
E[w(m)w^T(m)] = \tilde{Q}(m) \tag{5.14}
$$

$$
y(m) = C(m)x(m) + v(m) \tag{5.15}
$$

$$
E[v(m)v^T(m)] = R(m) \tag{5.16}
$$

Similarly, the Riccati equation, eq.'s(5.7,5.12), characterizes the error covariance for the optimal filter corresponding to the filtering problem involving the same state

equation, eq.(5.13,5.14), but with the following measurement equation.

$$\tilde{y}(m) = \begin{bmatrix} C(m) \\ D(m) \end{bmatrix} x(m) + u(m) \tag{5.17}$$

$$E[u(m)u^T(m)] = \begin{bmatrix} R(m) & 0 \\ 0 & I \end{bmatrix} \tag{5.18}$$

Since the filter corresponding to eq.(5.7,5.12) uses additional measurements compared to the filter corresponding to eq.(5.9,5.10), its error covariance can be no worse than the error covariance of the filter using fewer measurements; i.e. $P(m|m) \leq \overline{P}(m|m)$ or $\overline{P}^{-1}(m|m) \leq P^{-1}(m|m)$.

□

We now state and prove the following theorem concerning an upper bound for $P(m|m)$.

Theorem 5.1 *Given uniform upper boundedness of the stochastic reconstructibility grammian, i.e. eq.(5.4), and given uniform reconstructibility of the system there exists $\kappa > 0$ such that for all m at least M_0 levels from the initial level $P(m|m) \leq \kappa I$.*

Proof

Consider the following set of standard Riccati equations.

$$\overline{P}(m|m+1) = F(m+1)\overline{P}(m+1|m+1)F^T(m+1)$$
$$+ \quad G(m+1)\tilde{Q}(m+1)G^T(m+1) \tag{5.19}$$
$$\overline{P}^{-1}(m|m) = \overline{P}^{-1}(m|m+1) + C^T(m)R^{-1}(m)C(m) \tag{5.20}$$

From standard Kalman filtering results we know that given $(F(m), R^{-\frac{1}{2}}(m)C(m))$ is a uniformly observable pair that is bounded above, there exists a $\kappa > 0$ such that $\overline{P}(m|m) \leq \kappa I$ or $\overline{P}^{-1}(m|m) \geq \kappa^{-1}I$. But by Corollary 4.2, $(F(m), R^{-\frac{1}{2}}(m)C(m))$ being a uniformly observable pair is equivalent to the original system being uniformly reconstructible. Also, the grammian $(F(m), R^{-\frac{1}{2}}(m)C(m))$ being bounded above is equivalent to our assumption of uniform upper boundedness of the stochastic reconstructibility grammian. Thus, under uniform reconstructibility and the uniform upper

boundedness of the stochastic reconstructibility grammian of the original we deduce that $\overline{P}^{-1}(m|m) \geq \kappa^{-1}I$. But from Lemma 5.1 we know that $\overline{P}^{-1}(m|m) \leq P^{-1}(m|m)$. Thus, $P^{-1}(m|m) \geq \kappa^{-1}I$ or $P(m|m) \leq \kappa I$.

□

We can easily apply the previous ideas to derive an upper bound for $P_{ML}(m|m)$. Note that Lemma 5.1 would still apply if eq.(5.8) did not have the $P_x^{-1}(m)$ term; i.e. the lemma would apply to the case of the ML Riccati equations. Then by using the same argument used to prove Theorem 5.1 we can show the following theorem.

Theorem 5.2 *Given uniform upper boundedness of the stochastic reconstructibility grammian, i.e. eq.(5.4), and given uniform reconstructibility of the system there exists $\kappa' > 0$ such that for all m at least M_0 levels from the initial level $P_{ML}(m|m) \leq \kappa'I$.*

5.2 Lower Bound

We now derive a lower bound for $P(m|m)$. As in deriving the upper bound, we appeal heavily to standard system theory.

Lemma 5.2 *Let*

$$\overline{S}(m|m) \quad \triangleq \quad \frac{1}{2}(P^{-1}(m|m) - C^T(m)R^{-1}(m)C(m) + P_x^{-1}(m)) \quad (5.21)$$

$$\overline{S}(m|m+1) \quad \triangleq \quad F^{-T}(m+1)P^{-1}(m+1|m+1)F^{-1}(m+1) \quad (5.22)$$

Given the Riccati equation

$$S^*(m|m+1) \quad = \quad 2F^{-T}(m+1)S^*(m+1|m+1)F^{-1}(m+1)$$
$$+ \quad F^{-T}(m+1)C^T(m)R^{-1}(m)C(m)F^{-1}(m+1) \quad (5.23)$$
$$S^{*^{-1}}(m|m) \quad = \quad S^{*^{-1}}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1) \quad (5.24)$$

where $\overline{S}(0|0) = S^*(0|0)$. Then for all m $S^*(m|m) \geq \overline{S}(m|m)$.

Proof

By substituting eq.(5.12) into eq.(5.21) and collecting terms we get

$$\overline{S}(m|m) = P^{-1}(m|m+1) \tag{5.25}$$

By substituting eq.(3.1) into eq.(5.25) we arrive at

$$
\begin{aligned}
\overline{S}(m|m) &= [F(m+1)P(m+1|m+1)F^T(m+1) \\
&\quad + G(m+1)\tilde{Q}(m+1)G^T(m+1)]^{-1} \\
&= [\overline{S}^{-1}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1)]^{-1} \tag{5.26}
\end{aligned}
$$

where the the last equality results from the substitution of eq.(5.22). Also, by substituting eq.(5.21) into eq.(5.22) and collecting terms we get

$$
\begin{aligned}
\overline{S}(m|m+1) &= 2F^{-T}(m+1)\overline{S}(m+1|m+1)F^{-1}(m+1) \\
&\quad + F^{-T}(m+1)C^T(m)R^{-1}(m)C(m)F^{-1}(m+1) \\
&\quad - F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1) \tag{5.27}
\end{aligned}
$$

Now we prove by induction that for all m $S^*(m|m) \geq \overline{S}(m|m)$. Obviously, $S^*(0|0) \geq \overline{S}(0|0)$. As an induction hypothesis we assume $S^*(i+1|i+1) \geq \overline{S}(i+1|i+1)$. From eq.(5.27), eq.(5.23), and the fact that $F^{-T}(m+1)P_x^{-1}(m)F^{-1}(m+1) \geq 0$ we get that

$$S^{*^{-1}}(i|i+1) \leq \overline{S}^{-1}(i|i) \tag{5.28}$$

Substituting eq.(5.24) and eq.(5.26) into eq.(5.28) and cancelling terms we arrive at $S^{*^{-1}}(i|i) \leq \overline{S}^{-1}(i|i)$, i.e. $S^*(i|i) \geq \overline{S}(i|i)$.

□

Theorem 5.3 *Given uniform upper boundedness of the stochastic reachability grammian, i.e. eq.(5.3), and given uniform reachability of the system there exists $L > 0$ such that for all m at least M_0 levels from the initial level $P(m|m) \geq LI$.*

Proof

Consider the following set of standard Riccati equations.

$$
\begin{aligned}
S^*(m|m+1) &= 2F^{-T}(m+1)S^*(m+1|m+1)F^{-1}(m+1) \\
&\quad + F^{-T}(m+1)C^T(m)R^{-1}(m)C(m)F^{-1}(m+1) \qquad (5.29) \\
S^{*^{-1}}(m|m) &= S^{*^{-1}}(m|m+1) + G(m+1)\tilde{Q}(m+1)G^T(m+1) \qquad (5.30)
\end{aligned}
$$

From standard Kalman filtering results we know that if $(F^{-T}(m), G(m)\tilde{Q}^{\frac{1}{2}}(m))$ is a uniformly reachable pair that is bounded above, then there exists $N > 0$ such that $S^*(m|m) \leq NI$. However, from Corollary 4.1 and the invertibility of $F(m)$ the uniform reachability of the pair $(F^{-T}(m), G(m)\tilde{Q}^{\frac{1}{2}}(m))$ is equivalent to the original system being uniformly reachable. Also, the grammian $(F^{-T}(m), G(m)\tilde{Q}^{\frac{1}{2}}(m))$ being bounded above is equivalent to our assumption of uniform upper boundedness of the stochastic reachability grammian. Thus, under uniform reconstructibility and the uniform upper boundedness of the stochastic reconstructibility grammian of the original we deduce that $S^*(m|m) \leq NI$. But from Lemma 5.2 we know that $S^*(m|m) \geq \overline{S}(m|m)$. Thus, $\overline{S}(m|m) \leq NI$. But from eq.(5.21) we get

$$
\frac{1}{2}(P^{-1}(m|m) - C^T(m)R^{-1}(m)C(m) + P_x^{-1}(m)) \leq NI \qquad (5.31)
$$

It follows straightforwardly that

$$
P^{-1}(m|m) \leq L^{-1}I \qquad (5.32)
$$

where

$$
L^{-1}I \geq 2NI + C^T(m)R^{-1}(m)C(m) \qquad (5.33)
$$

Thus,

$$
P(m|m) \geq LI \qquad (5.34)
$$

$\square$

Using analagous arguments we can derive a lower bound for $P_{ML}(m|m)$. Note that with following definitions S^* obeys equations (5.23,5.24).

$$
S^*(m|m) \triangleq \frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \qquad (5.35)
$$

$$
S^*(m|m+1) \triangleq F^{-T}(m+1)P_{ML}^{-1}(m+1|m+1)F^{-1}(m+1) \qquad (5.36)
$$

Using the same argument as in the proof of Theorem 5.3 with our current definitions for S^* we get that

$$\frac{1}{2}(P_{ML}^{-1}(m|m) - C^T(m)R^{-1}(m)C(m)) \leq NI \qquad (5.37)$$

for $N > 0$. Equivalently,

$$P_{ML}^{-1}(m|m) \leq (L')^{-1}I \qquad (5.38)$$

for

$$(L')^{-1}I \geq 2NI + C^T(m)R^{-1}(m)C(m) \qquad (5.39)$$

Thus, we have the following theorem.

Theorem 5.4 *Given uniform upper boundedness of the stochastic reachability grammian, i.e. eq.(5.3), and given uniform reachability of the system there exists $L' > 0$ such that for all m $P_{ML}(m|m) \geq L'I$.*

6 Upward Stability on Trees

In this section we formalize the notion of stability for dynamic systems evolving up the tree. The dynamics on which we are interested in focusing the major portion of our analysis are the ML error dynamics of eq.(3.28). Thus the general class of systems we wish to study here has the form

$$z(t) = \mathcal{F}(m(t)+1)[z(\alpha t) + z(\beta t)] + \mathcal{G}(m(t))u(t) \tag{6.1}$$

What we wish to do is to study the asymptotic stability of this system as the dynamics propagate up the tree. Since we are interested in internal stability, we will consider the autonomous system with $u \equiv 0$.

Intuitively what we would like stability to mean is that $z(t) \to 0$ as we propagate farther and farther away from the initial level of the tree. Note, however, that as we move up the tree(or equivalently as the initial level moves farther down), $z(t)$ is influenced by a geometrically increasing number of nodes at the initial level. For example, $z(t)$ depends on $\{z(\alpha t), z(\beta t)\}$ or, alternatively on $\{z(\alpha^2 t), z(\beta\alpha t), z(\alpha\beta t), z(\beta^2 t)\}$ or, alternatively on $\{z(\alpha^3 t), z(\beta\alpha^2 t), z(\alpha\beta\alpha t), z(\beta^2\alpha t), z(\alpha^2\beta t), z(\beta\alpha\beta t), z(\alpha\beta^2 t), z(\beta^3 t)\}$, etc. Thus in order to study asymptotic stability it is **necessary** to consider an infinite dyadic tree, with an infinite set of initial conditions corresponding to all nodes at the initial level. Note also, that we might expect that there would be a number of meanings we could give to "$z(t) \to 0$" – e.g. do we consider individual nodes at a level or the infinite sequence of values at all points at a level?

To formalize the notion of stability let us change the sense of our index of recursion so that m increases as we move up the tree. Specifically, we arbitrarily choose a level of the tree to be our "initial" level, i.e. level 0, and we index the points on this initial level as $z_i(0)$ for $i \in \mathcal{Z}$. Points at the mth level up from level 0 are denoted $z_i(m)$ for $i \in \mathcal{Z}$. The dynamical equations we then wish to consider are of the form

$$z_i(m) = \mathcal{A}(m-1)(z_{2i}(m-1) + z_{2i+1}(m-1)) \tag{6.2}$$

Let $Z(m)$ denote the infinite sequence at level m, i.e. the set $\{z_i(m)\,,\, i \in \mathcal{Z}\}$.

The p-norm on such a sequence is defined as

$$\|Z(m)\|_p \triangleq (\sum_i \|z_i(m)\|_p^p)^{\frac{1}{p}} \qquad (6.3)$$

where $\|z_i(m)\|_p$ is the standard p-norm for the finite dimensional vector $z_i(m)$.

We define the following notion of exponential stability for a system.

Definition 6.1 *A system is l_p-exponentially stable if given any initial sequence $Z(0)$ such that $\|Z(0)\|_p < \infty$,*

$$\|Z(m)\|_p \leq C\alpha^m \|Z(0)\|_p \qquad (6.4)$$

where $0 \leq \alpha < 1$ and C is a positive constant.

From eq.(6.2) we can easily write down the following.

$$z_i(m) = \Phi(m, 0) \sum_{j \in O_{m,i}} z_j(0) \qquad (6.5)$$

where the cardinality of the set $O_{m,i}$ is 2^m and for $m_1 \geq m_2$

$$\Phi(m_1, m_2) = \begin{cases} I & m_1, m_2 \\ \mathcal{A}(m_1 - 1)\Phi(m_1 - 1, m_2) & m_1 > m_2 \end{cases} \qquad (6.6)$$

As in the case of standard dynamic systems it is the state transition matrix, $\Phi(m, 0)$, which plays a crucial role in studying stability on trees. However, unlike the standard case, as one can see from eq.(6.5), the nature of the initial condition that influences $z_i(m)$ depends crucially on m; in particular the number of points at level 0 to be summed up and scaled to give $z_i(m)$ is 2^m. These observations lead to the following:

Theorem 6.1 *The system defined in eq.(6.2) is l_p-exponentially stable if and only if*

$$2^{\frac{m(p-1)}{p}} \|\Phi(m, 0)\|_p \leq K'\gamma^m \quad \text{for all } m \qquad (6.7)$$

where $0 \leq \gamma < 1$ and K' is a positive constant.

Proof

Let us first show necessity. Specifically, suppose that for any $K > 0$, $0 \leq \gamma < 1$, and $M \geq 0$ we can find a vector z and an $m \geq M$ so that

$$\|\Phi(m,0)z\|_p > K\gamma^m 2^{-\frac{m}{q}}\|z\|_p \qquad (6.8)$$

where

$$\frac{1}{p} + \frac{1}{q} = 1 \qquad (6.9)$$

Let z and m be such a vector and integer for some choice of K, γ, and M, and define an initial sequence as follows. Let $\rho_0, \rho_1, \rho_2, \ldots$ be a sequence with

$$\sum_{i=0}^{\infty} \rho_i^p = 1 \qquad (6.10)$$

Then let

$$z_i(0) = \begin{cases} \rho_0 z & 0 \leq i < 2^m \\ \rho_1 z & 2^m \leq i < 2 \cdot 2^m \\ \vdots \\ \rho_i z & j2^m \leq i < (j+1)2^m \\ \vdots \end{cases} \qquad (6.11)$$

Note that

$$\begin{aligned} \|Z(0)\|_p^p &= \sum_{i=0}^{\infty} \|z_i(0)\|_p^p \\ &= 2^m \|z\|_p^p \end{aligned} \qquad (6.12)$$

Also, note that

$$\begin{aligned} z_i(m) &= \Phi(m,0) \sum_{j=i2^m}^{(i+1)2^m - 1} z_j(0) \\ &= 2^m \rho_i \Phi(m,0)z \end{aligned} \qquad (6.13)$$

Thus,

$$\begin{aligned} \|Z(m)\|_p^p &= 2^{mp} \|\Phi(m,0)z\|_p^p \\ &> 2^{mp} K^p \gamma^{mp} 2^{\frac{-mp}{q}} \|z\|_p^p \\ &= 2^{mp} K^p \gamma^{mp} 2^{\frac{-mp}{q}} 2^{-m} \|Z(0)\|_p^p \\ &= K^p \gamma^{mp} \|Z(0)\|_p^p \end{aligned} \qquad (6.14)$$

where the first equality comes from eq.(6.10), the inequality from eq.(6.8), the next equality from eq.(6.12), and the last equality from eq.(6.9). Hence for any K, $0 \leq \gamma < 1$ and $M \geq 0$ we can find an initial l_p-sequence $Z(0)$ and an $m \geq M$ so that

$$\|Z(m)\|_p > K\gamma^m \|Z(0)\|_p \tag{6.15}$$

so that the system cannot be l_p-exponentially stable.

To prove sufficiency we use the following.

Lemma 6.1 *A system is l_p-exponentially stable if for every i*

$$\|z_i(m)\|_p \leq K\beta^m \left(\sum_{j \in O_{m,i}} \|z_j(0)\|_p^p \right)^{\frac{1}{p}} \tag{6.16}$$

where $0 \leq \beta < 1$ and K is a positive constant.

Proof

By raising both sides of eq.(6.16) to the pth power we get

$$\|z_i(m)\|_p^p \leq K^p (\beta^p)^m \sum_{j \in O_{m,i}} \|z_j(0)\|_p^p \tag{6.17}$$

Since eq.(6.17) holds for every i we can write

$$\sum_i \|z_i(m)\|_p^p \leq K^p (\beta^p)^m \sum_i \|z_i(0)\|_p^p \tag{6.18}$$

The lemma follows from raising both sides of eq.(6.18) to the power of $\frac{1}{p}$.

$\square$

Lemma 6.2 *Consider the sequence of vectors x_i for $i \in \mathcal{Z}$. Then, for any m and any j*

$$\left\| \sum_{i \in O_{m,j}} x_i \right\|_p \leq 2^{\frac{m}{q}} \left(\sum_{i \in O_{m,j}} \|x_i\|_p^p \right)^{\frac{1}{p}} \tag{6.19}$$

where $O_{m,j} = \{j, j+1, ...j + 2^m - 1\}$ and q satisfies eq.(6.9).

Proof

We first show the following.

$$\|a + b\|_p \leq 2^{\frac{1}{q}}(\|a\|_p^p + \|b\|_p^p)^{\frac{1}{p}} \tag{6.20}$$

Since $\| \cdot \|_p^p$ is a convex function, we can write

$$\|(\frac{1}{2})a + (1 - \frac{1}{2})b\|_p^p \leq (\frac{1}{2})\|a\|_p^p + (1 - \frac{1}{2})\|b\|_p^p \tag{6.21}$$

from which eq.(6.20) follows immediately. We now show the result by induction on m. Suppose for all j

$$\|\sum_{i \in O_{m,j}} x_i\|_p \leq 2^{\frac{m}{q}}(\sum_{i \in O_{m,j}} \|x_i\|_p^p)^{\frac{1}{p}} \tag{6.22}$$

Consider the summing x_i over the two sets O_{m,j_1} and O_{m,j_2} where $j_2 = j_1 + 2^m$. From eq.(6.20) we get

$$\|(\sum_{i \in O_{m,j_1}} x_i + \sum_{i \in O_{m,j_2}} x_i)\|_p \leq 2^{\frac{1}{q}}(\|(\sum_{i \in O_{m,j_1}} x_i\|_p^p + \|(\sum_{i \in O_{m,j_2}} x_i\|_p^p)^{\frac{1}{p}} \tag{6.23}$$

Then by substituting into eq.(6.22) eq.(6.23) we get

$$\|\sum_{i \in O_{m,j_1} \cup O_{m,j_2}} x_i\|_p \leq 2^{\frac{(m+1)}{q}}(\|(\sum_{i \in O_{m,j_1}} x_i\|_p^p + \|(\sum_{i \in O_{m,j_2}} x_i\|_p^p)^{\frac{1}{p}} \tag{6.24}$$

□

We can now show sufficiency thereby completing the proof of the theorem. By applying the p-norm to eq.(6.5) and using the Cauchy-Schwarz inequality we get

$$\|z_i(m)\|_p \leq \|\Phi(m,0)\|_p\|\sum_{j \in O_{m,i}} z_j(0)\|_p \tag{6.25}$$

Using Lemma 6.2, we get

$$\|z_i(m)\|_p \leq \|\Phi(m,0)\|_p 2^{\frac{m}{q}}(\sum_{j \in O_{m,i}} \|z_j(0)\|_p^p)^{\frac{1}{p}} \tag{6.26}$$

By substituting eq.(6.7) into eq.(6.26) we get

$$\|z_i(m)\|_p \leq K'\gamma^m(\sum_{j \in O_{m,i}} \|z_j(0)\|_p^p)^{\frac{1}{p}} \tag{6.27}$$

which by Lemma 6.1 shows the system to be l_p-exponentially stable.

$\square$

Note that referring to eq.'s(6.2,6.5,6.6) we see that the l_p-exponential stability of eq.(6.2) is equivalent to the usual exponential stability of the system

$$\xi(m) = 2^{\frac{p-1}{p}} A(m-1)\xi(m-1) \tag{6.28}$$

For example for $p = 2$, we are interested in the exponential stability of

$$\xi(m) = \sqrt{2}A(m-1)\xi(m-1) \tag{6.29}$$

If A is constant this is equivalent to requiring A to have eigenvalues with magnitudes $< \frac{\sqrt{2}}{2}$.

Note also that it is straightforward to show that if one considers the system with inputs and outputs

$$
\begin{aligned}
z_i(m) &= \mathcal{A}(m-1)(z_{2i}(m-1) + z_{2i+1}(m-1)) \\
&\quad + \mathcal{B}(m-1)(u_{2i}(m-1) + u_{2i+1}(m-1)) \tag{6.30} \\
y_i(m) &= \mathcal{C}(m)z_i(m) \tag{6.31}
\end{aligned}
$$

then if $\mathcal{B}(m)$ and $\mathcal{C}(m)$ are bounded, the asymptotic stability of the undriven dynamics imply bounded-input/bounded-output stability.

7   Filter Stability

In this section we show that the error dynamics of the maximum likelihood filter are stable and also that the same is true of the overall filter.

Theorem 7.1 *Suppose that the system is uniformly reachable and uniformly reconstructible. Then, the error dynamics of the maximum likelihood filter are l_2-exponentially stable.*

Proof

The following proof follows closely the standard proof for stability of discrete-time Kalman filters given in [9]. Based on the comments at the end of the preceding section and on the ML error dynamics of eq.(3.28), we see that we wish to show that the following causal system is stable in the standard sense.

$$z(m) = P_{ML}(m|m)P_{ML}^{-1}(m|m-1)\sqrt{2}F(m-1)z(m-1) \qquad (7.1)$$

Theorem's 5.2 and 5.4, i.e. the upper and lower bounds on $P_{ML}(m|m)$, allow us to define the following Lyapunov function.

$$V(z,m) \overset{\triangle}{=} z^T(m)P_{ML}^{-1}(m|m)z(m) \qquad (7.2)$$

Let us also define the following quantity.

$$\tilde{z}(m) \overset{\triangle}{=} \sqrt{2}F(m-1)z(m-1) \qquad (7.3)$$

$$= P_{ML}(m|m-1)P_{ML}^{-1}(m|m)z(m) \qquad (7.4)$$

Substituting eq.(3.7) into eq.(7.2) followed by algebraic manipulations, one gets

$$
\begin{aligned}
V(z,m) &= z^T(m)(2P_{ML}^{-1}(m|m-1) + C^T(m)R^{-1}(m)C(m))z(m) && (7.5)\\
&= 2z^T(m)(P_{ML}^{-1}(m|m) - 2P_{ML}^{-1}(m|m-1))z(m) - z^T(m)C^T(m)R^{-1}(m)C(m))z(m)\\
&+ z^T(m)(2P_{ML}^{-1}(m|m-1))z(m)\\
&+ \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m-1)\frac{\tilde{z}(m)}{\sqrt{2}} - \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m-1)\frac{\tilde{z}(m)}{\sqrt{2}} && (7.6)\\
&= -(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})^T P_{ML}^{-1}(m|m-1)(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})\\
&- z^T(m)C^T(m)R^{-1}(m)C(m)z(m) + \frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m-1)\frac{\tilde{z}(m)}{\sqrt{2}} && (7.7)
\end{aligned}
$$

But note that by using the matrix inversion lemma we get

$$\frac{\tilde{z}^T(m)}{\sqrt{2}}P_{ML}^{-1}(m|m-1)\frac{\tilde{z}(m)}{\sqrt{2}} = V(z,m-1) - \Delta \qquad (7.8)$$

$$\Delta \geq 0 \qquad (7.9)$$

It follows that

$$
\begin{aligned}
V(z,m) - V(z,m-1) &\leq -(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})^T P_{ML}^{-1}(m|m-1)(\sqrt{2}z(m) - \frac{\tilde{z}(m)}{\sqrt{2}})\\
&- z^T(m)C^T(m)R^{-1}(m)C(m)z(m) && (7.10)
\end{aligned}
$$

Stability follows from eq.(7.10) under the condition of uniform observability of the pair $(F(m), R^{-\frac{1}{2}}(m))C(m)$ which by Corollary 4.2 is equivalent to uniform reconstructibility of the system.

$\square$

Let us now examine the full estimation error after incorporating prior statistics. It is straightforward to see that

$$\tilde{x}(t|t) = P(m(t)|m(t))(P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t) + P_x^{-1}(m(t))x(t)) \qquad (7.11)$$

Thus we can view $\tilde{x}(t|t)$ as a linear combination of the states of two upward-evolving systems, eq.(3.28) for $\tilde{x}_{ML}(t|t)$ and one for $P_x^{-1}(m(t))x(t)$. Note first that since $P(m|m) \leq P_{ML}(m|m)$

$$\|P(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t))\tilde{x}_{ML}(t|t)\| \leq \|\tilde{x}_{ML}(t|t)\| \qquad (7.12)$$

and we already have the stability of the $\tilde{x}_{ML}(t|t)$ dynamics from Theorem 7.1. Turning to the second term in eq.(7.11), note first that thanks to Theorem 5.1, $P(m(t)|m(t))$ is bounded. Note also that the covariance of $P_x^{-1}(m(t))x(t)$ is simply $P_x^{-1}(m(t))$. By uniform reachability $P_x^{-1}(m(t))$ is bounded above. Thus, while $P_x(m(t))$ might diverge, the contribution to the error of the second term in eq.(7.11) is bounded.

Also, our previous analysis allows us to conclude that the full, driven $\tilde{x}_{ML}(t|t)$ dynamics are bounded-input, bounded-output stable from inputs $\tilde{w}$ and v to output $\tilde{x}_{ML}(t|t)$. If we use eq.(3.19), together with eq.(2.3) and eq.'s(2.6-2.8) we can write down the following upward dynamics for $\xi(t) = P_x^{-1}(m(t))x(t)$:

$$\begin{aligned} \xi(t) &= \frac{1}{2}A^T(m(t)+1)(\xi(\alpha t) + \xi(\beta t)) \\ &+ \frac{1}{2}N(m(t)+1)(\tilde{w}(\alpha t) + \tilde{w}(\beta t)) \end{aligned} \qquad (7.13)$$

where

$$N(m(t)+1) = P_x^{-1}(m(t))A^{-1}(m(t)+1)B(m(t)+1) \qquad (7.14)$$

Note that in general there is no reason to constrain the autonomous dynamics of eq.(7.13) to be stable. However, if they are not, then reachability implies that $P_x(m) \to \infty$ so that $N(m) \to 0$ and the covariance of $\tilde{w} \to I$. The bounded-input, bounded-output stability of this system can be easily checked.

8 Steady-state Filter

In this section we study properties of our filter under steady-state conditions; i.e. we analyze the asymptotic properties of the filter. We state and prove several results. First we show that the error covariance of the ML estimator converges to a steady-state limit and that furthermore, the steady-state filter is l_2-exponentially stable.

Theorem 8.1 *Consider the following system defined on a tree.*

$$x(t) = \frac{1}{2}F(x(\alpha t) + x(\beta t)) + \frac{1}{2}G(\tilde{w}(\alpha t) + \tilde{w}(\beta t)) \qquad (8.1)$$

$$y(t) = Cx(t) + v(t) \qquad (8.2)$$

$$E[\tilde{w}(t)\tilde{w}^T(t)] = \tilde{Q} \qquad (8.3)$$

$$E[v(t)v^T(t)] = R \qquad (8.4)$$

where $v(t)$ is white and $w(t)$ is white in subtrees. Suppose that $(F, G\tilde{Q}^{\frac{1}{2}})$ is a reachable pair and $(F, R^{-\frac{1}{2}}C)$ is an observable pair. The error covariance for the ML estimator, $P_{ML}(m|m)$, converges as $m \rightarrow \infty$ to $\overline{P}_\infty$, which is the unique positive definite solution to

$$\begin{aligned}\overline{P}_\infty &= \frac{1}{2}F\overline{P}_\infty F^T + \frac{1}{2}G\tilde{Q}G^T \\ &- K_\infty(\frac{1}{2}CF\overline{P}_\infty F^T C^T + \frac{1}{2}CG\tilde{Q}G^T C^T + R)K_\infty^T \qquad (8.5)\end{aligned}$$

where

$$K_\infty = \overline{P}_\infty C^T R^{-1} \qquad (8.6)$$

Moreover, the autonomous dynamics of the steady-state ML filter, i.e.

$$e(t) = \frac{1}{2}(I - K_\infty C)F(e(\alpha t) + e(\beta t)) \qquad (8.7)$$

are l_2-exponentially stable.

Proof

Recall the Riccati equations for the ML estimator where the scale variable m increases in the direction upward along the tree.

$$P_{ML}(m|m+1) = FP_{ML}(m+1|m+1)F^T + G\tilde{Q}G^T$$

$$(8.8)$$

$$P_{ML}^{-1}(m|m) = 2P_{ML}^{-1}(m|m+1) + C^T R^{-1} C \qquad (8.9)$$

Convergence of $P_{ML}(m|m)$

In order to show the existence of a limit of $P_{ML}(m|m)$ as $m \to \infty$ we show that both a) $P_{ML}(m|m)$ is monotone-nonincreasing in m and b) $P_{ML}(m|m)$ is bounded below.

a) We adopt the following notation.

$$P(m) \triangleq P_{ML}(m|m) \quad m \geq 0 \qquad (8.10)$$

$$P(m;m') \triangleq P(m-m') \quad m \geq m' \qquad (8.11)$$

By the scale-invariance of our system showing

$$m_1 < m_2 \to P(m;m_1) \leq P(m;m_2) \qquad (8.12)$$

is equivalent to demonstrating that $P(m)$ is monotone-nonincreasing.

We note that eq.'s(8.8,8.9) preserve positive definite orderings; i.e. if $P_1(m_2) \leq P_2(m_2)$ then $P_1(m;m_2) \leq P_2(m;m_2)$ for $m \geq m_2$. We now take

$$P_1(m_2) = P(m_2;m_1) \qquad (8.13)$$

$$P_2(m_2) = \infty \text{ (initial condition for the ML estimator)} \qquad (8.14)$$

Then,

$$P_1(m;m_2) = P(m;m_1) \qquad (8.15)$$

$$P_2(m;m_2) = P(m;m_2) \qquad (8.16)$$

for $m \geq m_2$. So by the property of postitive definite ordering of the Riccati equations we know that

$$P_1(m;m_2) \leq P_1(m;m_2) \qquad (8.17)$$

and thus,

$$P(m; m_1) \leq P(m; m_2) \tag{8.18}$$

b) The fact that $P_{ML}(m|m)$ is bounded below follows from Theorem 5.4 under our assumptions of reachability and observability.

Having established the convergence of $P_{ML}(m|m)$, let us denote the limit as follows.

$$\lim_{m \to \infty} P_{ML}(m|m) \stackrel{\triangle}{=} \overline{P}_\infty \tag{8.19}$$

Note that by Theorem 5.4 $\overline{P}_\infty$ must be positive definite. We can also establish that $P_{ML}(m|m)$ must converge to the solution of the steady state Riccati eq.(8.5). Since $P_{ML}(m|m)$ both satisfies the Riccati eq.'s(8.8,8.9) and converges to a limit, this limit must satisfy the fixed point equation for eq.'s(8.8,8.9). This fixed point equation is precisely the steady state Riccati eq.(8.5).

Exponential Stability of $\frac{1}{2}(I - K_\infty C)F$

In order for $\frac{1}{2}(I - K_\infty C)F$ to be l_2-exponentially stable, it must have eigenvalues that are strictly less than $\frac{\sqrt{2}}{2}$. This fact follows from Theorem 6.1.

From Theorem 7.1 we know that the following system is exponentially stable with respect to $\| \cdot \|_2^*$.

$$z(t) = P_{ML}(m(t)|m(t))P_{ML}^{-1}(m(t)|m(t) - 1)(z(\alpha t) + z(\beta t)) \tag{8.20}$$

which can be rewritten as

$$z(t) = \frac{1}{2}(I - K(m(t))C)F(z(\alpha t) + z(\beta t)) \tag{8.21}$$

where

$$K(m(t)) = P_{ML}(m(t)|m(t))C^T R^{-1} \tag{8.22}$$

But, since $\lim_{m \to \infty} P_{ML}(m|m) = \overline{P}_\infty$, the system in eq.(8.21) in steady-state becomes

$$z(t) = \frac{1}{2}(I - K_\infty C)F(z(\alpha t) + z(\beta t)) \tag{8.23}$$

Uniqueness of $\overline{P}_\infty$

Consider P_1 and P_2, both of which satisfy the steady state Riccati eq.(8.5). Thus,

$$P_1 = \frac{1}{2}FP_1F^T + \frac{1}{2}G\tilde{Q}G^T$$
$$- K_1(\frac{1}{2}CFP_1F^TC^T + \frac{1}{2}CG\tilde{Q}G^TC^T + R)K_1^T \qquad (8.24)$$

$$P_2 = \frac{1}{2}FP_2F^T + \frac{1}{2}G\tilde{Q}G^T$$
$$- K_2(\frac{1}{2}CFP_2F^TC^T + \frac{1}{2}CG\tilde{Q}G^TC^T + R)K_2^T \qquad (8.25)$$

Subtracting eq.(8.25) from eq.(8.24) we get

$$P_1 - P_2 = \frac{\sqrt{2}}{2}(I - K_1C)F(P_1 - P_2)(\frac{\sqrt{2}}{2}(I - K_1C)F)^T$$
$$+ \Delta \qquad (8.26)$$

where Δ is a symmetric matrix. Note that we have established the fact that $\frac{\sqrt{2}}{2}(I - K_1C)F$ has eigenvalues within the unit circle. From standard system theory this tells us that we can write $P_1 - P_2$ as a sum of positive semidefinite terms. This implies that $P_1 - P_2$ is positive semidefinite or $P_1 \geq P_2$. By subtracting eq.(8.24) from eq.(8.25) and using the same argument we can establish that $P_2 \geq P_1$.

□

Note that the preceding analysis assumed constant matrices $F, G, C, \tilde{Q}$, and R. If we begin with our original downward model eq.(2.2), eq.(2.9) with A, B, C, Q, and R invertible, the constancy of F, G, and $\tilde{Q}$ require that P_x^{-1} is constant. As we are interested in asymptotic behavior, there is no loss of generality in assuming this and there are two distinct cases. Specifically, if A is stable, then the covariance $P_x(m(t))$ at all finite nodes(starting from an infinitely remote coarse level) is the positive definite(because of reachability) solution P_x of eq.(2.4), and in this case, we have that

$$P(m|m) \rightarrow (\overline{P}_\infty^{-1} + P_x^{-1})^{-1} \qquad (8.27)$$

On the other hand, if A is unstable, $P_x^{-1}(m(t)) \rightarrow 0$ and

$$P(m|m) \rightarrow \overline{P}_\infty \qquad (8.28)$$

Note that the existence of two distinct limiting forms for $P(m|m)$, depending on the stability of the original model is another significant deviation from standard causal theory.

9 Summary

In this paper we have analyzed in detail the filtering step of the Rauch-Tung-Striebel smoothing algorithm developed in [7] for the optimal estimation of a class of multiresolution stochastic processes. In particular we have developed the system-theoretic concepts necessary for the analysis of the stability and the steady-state properties of the filter. Notions of stability, reachability, and observability were developed for systems whose dynamics evolve upward on a dyadic tree. We then used these notions in showing stability of the optimal filter and steady-state convergence of the filter.

References

[1] B. Anderson and T. Kailath, "Forwards, backwards, and dynamically reversible Markovian models of second-order processes," *IEEE Trans. Circuits and Systems*, CAS-26, no. 11, pp. 956–965, 1978.

[2] M. Basseville, A. Benveniste, A.S. Willsky, and K.C. Chou, "Multiscale Statistical Processing: Stochastic Processes Indexed by Trees," in *Proc. of Int'l Symp. on Math. Theory of Networks and Systems*, Amsterdam, June 1989.

[3] A. Brandt, "Multi-level adaptive solutions to boundary value problems," *Math. Comp.*, vol. 13, pp. 333–390, 1977.

[4] W. Briggs, *A Multigrid Tutorial*, Philadephia: SIAM, 1987.

[5] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Comm.*, vol. 31, pp. 482–540, 1983.

[6] K.C. Chou, *A Stochastic Modeling Approach to Multiscale Signal Processing*, MIT, Department of Electrical Engineering and Computer Science, Ph.D. Thesis, (in preparation).

[7] K.C. CHOU, A.S. WILLSKY, A. BENVENISTE, AND M. BASSEVILLE, "Recursive and Iterative Estimation Algorithms for Multi-Resolution Stochastic Processes," Proc. 28th IEEE Conf. on Dec. and Cont., Tampa, Dec. 1989.

[8] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. on Pure and Applied Math.*, vol. 91, pp. 909– 996, 1988.

[9] J. Deyst, C. Price, "Conditions for Asymptotic Stability of the Discrete, Minimum Variance, Linear, Estimator," *IEEE Trans. on Automatic Control*, vol. 13, pp. 702– 705, Dec. 1968.

[10] S.G. MALLAT, "A theory for multiresolution signal decomposition: the wavelet representation," Dept. of Computer and Info. Science—U. of Penn., MS–CIS–87–22, GRASP LAB 103, May 1987.

[11] S.G. MALLAT, "Multiresolution approximation and wavelets," Dept. of Computer and Info. Science—U. of Penn., MS–CIS–87–87, GRASP LAB 80, Sept. 1987.

[12] T. Verghese and T. Kailath, "A further note on backward Markovian models," *IEEE Trans. on Information Theory*, IT–25, pp. 121–124, 1979.

We define the following quantities.

$$Y_{M,t_0} = \mathcal{H}_M X_{M,t_0} \tag{.29}$$

$$x_{t_0} = \Phi(t_0) X_{M,t_0} \tag{.30}$$

$$\Phi(t_0) = G I_{2^M} \tag{.31}$$

where G is invertible(and thus $\Phi(t_0)$ is onto). We use $\mathcal{N}(\cdot)$ and $\mathcal{R}(\cdot)$ to denote nullspace and rangespace, respectively. A system is upward-reconstructible if given Y_{M,t_0}, x_{t_0} is uniquely determined, i.e. $\mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0))$. We first prove the following lemma.

Lemma .1 *For all M*

$$\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0) = \Lambda \Phi^T(t_0) \tag{.32}$$

where

$$\Lambda = diag(\ \underbrace{\overline{\Lambda}...\overline{\Lambda}}_{2^M\ times}\ ) \tag{.33}$$

and $\overline{\Lambda}$ is some matrix.

Proof

The structure of $\mathcal{H}_M^T \mathcal{H}_M$, which we denoted as $\mathcal{M}_M$, is described in a recursive fashion in eq.'s(4.20-4.24). We compute

$$
\begin{aligned}
\mathcal{M}_M \Phi^T(t_0) &= U(M,0)\Phi^T(t_0) \\
&= \begin{bmatrix} U(M,1)G^T I_{2^{M-1}}^T + 2^{M-1}T(M,0)G^T I_{2^{M-1}} \\ 2^{M-1}T(M,0)G^T I_{2^{M-1}} + U(M,1)G^T I_{2^{M-1}}^T \end{bmatrix}
\end{aligned} \tag{.34}
$$

By repeating this procedure $M-1$ more times we get

$$U(M,0)\Phi^T(t_0) = \Lambda \Phi^T(t_0) \tag{.35}$$

where

$$\Lambda = diag(\ \underbrace{\overline{\Lambda}...\overline{\Lambda}}_{2^M\ times}\ ) \tag{.36}$$

and

$$\overline{\Lambda} = \sum_{i=0}^{M-1} 2^{M-1-i} T(M,i) + U(M,M) \qquad (.37)$$

□

We prove the following theorem.

Theorem .1 $\mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0))$ *iff* $\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0)$ *is invertible.*

Proof

a)

$$\mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0)) \longrightarrow \Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0) \text{ is invertible}$$

Assume $\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0)$ is not invertible. Then for some $y \neq 0$, $y^T\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0)y = 0$. This implies $\mathcal{H}_M \Phi^T(t_0)y = 0$. But the fact that $\Phi(t_0)$ is onto implies $\Phi^T(t_0)y \neq 0$. Furthermore, $\Phi^T(t_0)y \neq 0$ implies $\Phi(t_0)\Phi^T(t_0)y \neq 0$ since if it were true that $\Phi(t_0)\Phi^T(t_0)y = 0$, then $y^T\Phi(t_0)\Phi^T(t_0)y = 0$, which implies $\Phi^T(t_0)y = 0$. Thus, there exists a $z \neq 0$, namely $\Phi^T(t_0)y$, such that $\mathcal{H}_M z = 0$ and $\Phi(t_0) \neq 0$; i.e. it is not true that $\mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0))$.

b)

$$\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0) \text{ is invertible} \longrightarrow \mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0))$$

Assume that $\mathcal{N}(\mathcal{H}_M) \subset \mathcal{N}(\Phi(t_0))$ is false; i.e. there exists an x such that $\mathcal{H}_M x = 0$ and $\Phi(t_0)x \neq 0$. Since $x \in \mathcal{R}(\Phi^T(t_0)) \oplus \mathcal{N}(\Phi(t_0))$, we can write $x = x_{\mathcal{R}(\Phi^T(t_0))} + x_{\mathcal{N}(\Phi(t_0))}$ where $x_{\mathcal{R}(\Phi^T(t_0))}$ is non-zero and $x_{\mathcal{N}(\Phi(t_0))}$ may or may not be non-zero. Since $\mathcal{H}_M x = 0$, $\mathcal{H}_M x_{\mathcal{R}(\Phi^T(t_0))} + \mathcal{H}_M x_{\mathcal{N}(\Phi(t_0))} = 0$, which means that $\mathcal{H}_M \Phi^T(t_0)y + \mathcal{H}_M x_{\mathcal{N}(\Phi(t_0))} = 0$ for some $y \neq 0$. Left multiplying by $\Phi(t_0)\mathcal{H}_M^T$, we get

$$\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M \Phi^T(t_0)y + \Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M x_{\mathcal{N}(\Phi(t_0))} = 0 \qquad (.38)$$

But from Lemma .1 and our definition for $\Phi(t_0)$, we get

$$\Phi(t_0)\mathcal{H}_M^T \mathcal{H}_M = \Phi(t_0)\Lambda^T = G\overline{\Lambda}^T [\ \underbrace{I...I}\ ] \qquad (.39)$$
$$2^M \text{ times}$$

By substituting (.39) into (.38), we get

$$\Phi(t_0)\mathcal{H}_M^T\mathcal{H}_M\Phi^T(t_0)y + G\overline{\Lambda}^T[\underbrace{I...I}_{2^M \text{ times}}]x_{\mathcal{N}(\Phi(t_0))} = 0 \qquad (.40)$$

for $y \neq 0$. But $x_{\mathcal{N}(\Phi(t_0))} \in \mathcal{N}(\Phi(t_0))$ implies that $\Phi(t_0)x_{\mathcal{N}(\Phi(t_0))} = 0$ or, using the definition of $\Phi(t_0)$, $G[\underbrace{I...I}_{2^M \text{ times}}]x_{\mathcal{N}(\Phi(t_0))} = 0$. But since G is invertible, then $[\underbrace{I...I}_{2^M \text{ times}}]x_{\mathcal{N}(\Phi(t_0))} = 0$. Thus, eq.(.40) collapses to $\Phi(t_0)\mathcal{H}_M^T\mathcal{H}_M\Phi^T(t_0)y = 0$ for some $y \neq 0$, implying that $y^T\Phi(t_0)\mathcal{H}_M^T\mathcal{H}_M\Phi^T(t_0)y = 0$ for some $y \neq 0$; i.e. $\Phi(t_0)\mathcal{H}_M^T\mathcal{H}_M\Phi^T(t_0)$ is not invertible.

□